Behavior Research Methods & Instrumentation
1979, Vol. 11 (2), 172-176

The computer and eye processing pictures: The implementation of a raster graphics device

MARCEL ADAM JUST and PATRICIA CARPENTER
Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

The process of solving visual analogies was examined using a digital video display and video-based eye tracker. The results indicated that increasing the complexity of an analogy by increasing the number of relations involved in the problem resulted primarily in increasing the time it takes to discover the mapping between the two members of the first pair. The display of the picture and the analysis of the eye fixations are described.

Picture processing is reemerging as an important area of research in psychology. The reasons for its neglect are primarily theoretical. Unlike verbal information, images are difficult to represent in a psychological theory. One approach has been to utilize verbal-based theories. Unfortunately, the translation of verbal theories into the pictorial domain has not always been successful. For example, some of the earlier discussions about "picture grammars" have not yielded general grammars for scenes that are more complex than block figures. Interestingly, the theoretical problem parallels a technological one. It has been difficult to represent pictures in a digital computer. This complicates the presentation and analysis of pictures. This problem is particularly crucial in eye fixation research where theoretical issues require an analysis of what is examined in the picture.

There have been three typical solutions to the problem of displaying pictures in an experiment. One is to present the actual image, or to present it through some off-line device, such as a slide projector or a closed-circuit television. This presentation mode has problems when it is important to know the coordinates of various components of the display. To analyze eye fixation data, it is important that the coordinates of various components be kept identical within the experiment. A slight jarring of the pictures, projector or camera causes misalignment. Thus, the off-line presentation approach presents calibration problems that can be extremely time consuming.

A second approach is to use a vector graphics device, such as the DEC VT-11. The graphics device is best suited for presenting simple line drawings. The strength of the vector graphics is programmability. The weakness is that the device displays only lines; it cannot be used for shading or natural pictures of any complexity. Even a detailed line drawing may tax the refresh capability of the display. In addition, there may be a problem inputting image information. Specifying the coordinates of irregular lines may be impractical. A light pen can be used to draw a picture of moderate complexity, but it lacks precision.

A third approach is to use a raster graphics device. The hardware part of the device described here was designed for our laboratory by Powell (1978). The device, a digital video controller, creates a digital representation of an arbitrary input to an ordinary video camera. Figure 1 shows the original black-and-white drawing. Figure 2 shows the form of the image displayed on a video monitor. The controller can also be used to digitize pictures with multiple gray levels. The digital representation is then stored in a disk file; it can be accessed during an experiment, undigitized by the controller, and output to an ordinary video monitor that the subject views. Obviously, this digital video controller solves the input problem: It is no more difficult to input a complex picture than a simple one. It also solves the calibration problem, since the coordinates of the pictures are under program control.

In this paper we discuss the video digital controller as one solution to the problem of picture presentation and analysis. We present an application of the controller in an experiment on eye fixations during the solution of visual analogies to give an idea of its actual role in research. Then, we describe the controller's software support in more detail.

THE APPLICATION: THE RAVEN PROGRESSIVE MATRICES TEST

For some time, we have been interested in how people perform the Raven advanced progressive matrices, a set of visual analogies (Raven, 1962/1974). Figure 1 shows the format of a Raven-type problem. (We constructed this problem; Raven items are used in the experiments.) The subject's task is to choose the correct alternative that completes the 3 by 3 analogy. There are 36 problems arranged in order of increasing difficulty. The Raven is an interesting domain in which to examine visual problem solving. In addition,

The order of the authors is arbitrary. This research was supported by the National Institute of Mental Health Grant 29617-02 and the National Institute of Education Grant 77007.

0005-7878/79/020172-05$00.75/0

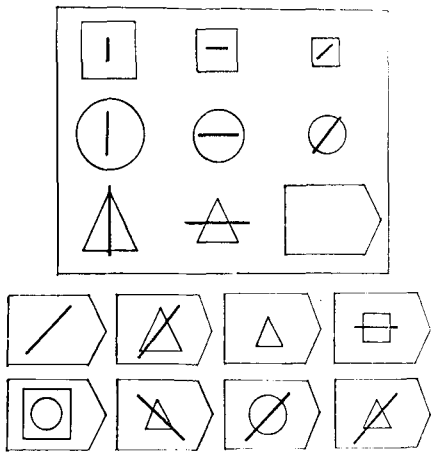

Figure 1. Drawing of a Raven-type problem that was digitized by the digital video controller.

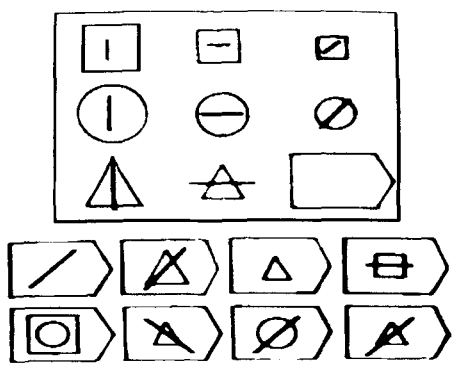

Figure 2. Photograph of a monitor showing the digitized display of the drawing in Figure 1.

it is used as a nonverbal intelligence test. In view of the practical role it plays in education, it is appropriate to analyze the kinds of processes it taps (cf. Hunt, 1974).

We were interested in whether certain structural variables could account for performance in the Raven. The two structural variables were the number of transformations relating the elements in a row or column and the number of elements in a problem. For example, there are two elements in each figure of the array shown in Figure 2, a line and a geometric shape. There are two transformations relating each pair of figures. Across the rows, the line changes orientation and the shape changes size. Down a column, the line changes size and the geometric shape changes from square to circle to triangle. Typically, the number of transformations between row elements is the same as the number between column elements.

These two structural variables accounted for approximately 50% of the variance among the 36 problems for a set of normative error data collected by Forbes (1964) from 2,256 subjects, and for a similar proportion of the variance for a set of reaction times collected from 11 Carnegie-Mellon University undergraduates. Thus, these two factors play important roles in the Ravens, although other factors, including the kind of transformations and the particular values of the transformation, play a role in the remaining variance. The number of transformations accounts for about twice as much variance as does the number of elements.

We have begun two experimental investigations of the processing implications of these two variables. In one, we are examining how people solve the Raven test by recording their eye fixations during problem solving; in the other, we constructed 28 2 by 2 analogies that use the same kinds of transformations and elements as the Raven (see also Sternberg, 1977; Pellegrino & Glaser, Note 1). We experimentally manipulated the number of transformations and elements and examined the subjects' eye fixations while they decided whether the bottom right-hand figure of the 2 by 2 analogy correctly or incorrectly completed the analogy.

Method

In both experiments, we digitized the problems using the digital video controller. The 2 by 2 analogies were produced by manual drawing. For the Raven analogies, the actual test items were used. The image was input by pointing a camera at the drawing, digitizing the video signal, and storing the digitized representation. During the experiment, this representation was undigitized by the controller and output to a video monitor that the subject viewed. The problems were presented using a PDP-11/04. The subject's eye fixations were tracked, and the coordinates of the eye's point of regard were recorded and stored for later analysis. The subject's eye fixations were calibrated so that the locus of the eye fixations could be scored with respect to the elements in a problem.

The digitization was done during two "photographic" sessions. The particular stimuli used in these experiments were digitized using a single-bit mode so that the image was black on a white blackground. By appropriately lighting the drawing and working with the camera, it is possible to get a good representation of most of the Raven problems. Very fine lines tax the resolution of the video controller. A file editor allows one to later delete or amend the file. There is often noise in an individual image caused by uneven lighting or inaccuracy in the A/D convertor. This noise can be corrected using an editor that will change the individual elements, pixels, that make up the picture. During the experiment, the experimental program can access the file that contains the pictures and present them either in random order or in an order specified by the experimenter. These programs are described in more detail later.

Eye tracking is done by an Applied Science Laboratories eye view monitor, a video-based system that monitors corneal and pupil reflection and computes the point of regard from the two reference points. Both the eye tracker and the video controller operate at 60 Hz; in fact, they are both driven by the video controller's synch pulse. We also have a videotaped record of each trial consisting of cross-hairs indicating the point of regard superimposed on the scene the subject is viewing.

The output of the tracker is the X and Y coordinates of the subject's point of fixation every 16.7 msec. These coordinates are input to the PDP-11/04 for a limited amount of on-line processing. The on-line processing program aggregates the coordinates into fixations. The parameters defining a fixation are set by the experimenter.

The later data analysis aggregates the fixations into sectors that correspond to the elements in a picture. For example, in the experiment involving 2 by 2 analogy problems, the program computed the order in which the four cells of the matrix were fixated and the duration on each cell. For convenience, the two cells in the top row are called A and B, and the two bottom cells C and D. A typical output might indicate that the subject looked at A for 500 msec, followed by B for 300 msec, and so forth. Further analyses can be done on the particular sequential and temporal aspects of the eye fixations.

Results

In a number of domains involving visual stimuli, such as mental rotation and reading, the locus and duration of eye fixations are sensitive to the mental processes (Carpenter & Just, 1977; Just & Carpenter, 1976a, 1976b). Similarly, in these Raven problems eye fixations index the sequence of problem solving stages. So far the analysis is limited to five people selected for being "good" on these kinds of problems. First, the subjects extract a great deal of information from the 3 by 3 matrix before looking at the eight answer alternatives. Second, if the problem is easy (normatively or structurally defined), they may use either a row or a column organization in scanning, but not both. Third, there are distinct stages in the problem solution process. One figure is initially encoded and then compared to a second. During the comparison process, subjects may begin to abstract what dimension is varied. Usually, they scan an entire row before stating a hypothesized dimension (e.g., orientation). They may then go back and encode specific values along the dimension, and they may then state a rule and test it on a row they have encoded or on another row. If the problem is particularly easy, with few transformations, they do not fixate all of the elements in the 3 by 3 array.

THE ANALOGY TASK

The 2 by 2 true-false analogies show important similarities to the actual Raven. Again, the number of transformations accounts for more of the variance among the solution times than does the number of elements. This suggests that elements that are not transformed do not play a large role in the processing. Once the subject determines that an element is not being transformed, it becomes an irrelevant dimension. The analysis of the sequence of eye fixations indicated that the number of transformations had a large effect on how the subject encoded A and B. The analysis showed that subjects looked at A and then crossed to B and then often back to A. Subjects spent more time abstracting the relations between A and B than applying them to C and D. There were more crossings between A and B (3.6) than between C and D (2.9). There was on the average only one crossing from A to C and from B to D. As the number of transformations increased, the number of crossings increased, and most of the increase was due to an increased number of crossings from A to B.

The eye fixations suggest that subjects encode A and, by comparing it to B, extract the relevant dimensions and the precise values of A and B. The initial crossings between A and B may also be used to hypothesize the rule that is then applied to C. The crossings between C and D reflect the application of the rule and the verification of whether D has the correct value.

The eye fixations also indicate that our subjects tended to have difficulty with more than two rules. A problem with one or two rules would have a sequence of fixations between A and B and then between C and D. If there were three transformations, subjects were more likely to repeat the cycle, presumably picking up the next transformation.

The analysis of fixation durations also supported the hypothesis that the most time-consuming aspect of the problem was in abstracting the rule. Subjects spent most of their time on A (30%). As the number of transformations increased, subjects spent progressively more time determining the relation between A and B and less time looking at D.

THE DIGITAL VIDEO CONTROLLER

The video controller can digitize in four different modes. It can produce a 16-level gray scale representation that looks fairly natural or a two-level (single-bit) representation that looks much like a Kodalith print. In addition, it can digitize at full resolution (256,000 pixels) or at half resolution. The largest picture that can be produced is constrained by the 16,000-word local memory of the video controller. Any picture in a file on a disk can be randomly accessed, read into memory, and displayed in less than 500 msec by means of a direct memory access (DMA) transfer. The hardware aspects of the controller are described in Powell (1978). We next describe the software created to implement the video controller.[1]

Several kinds of software support are needed to efficiently use the digital video controller. For instance, software is needed for digitizing a picture, for presenting the picture, for creating a picture, for cleaning up a picture that has noise, and for presenting alphanumerics. These functions are handled by four programs: (1) an editor for file handling and writing, (2) a picture presentation program, (3) a picture editor, and (4) an alphanumerics presentation program.

The editor allows one to write new files of pictures, to read existing files, or to read from and write to existing files. This is just an ordinary file editor, similar to the RT-11 editor. With the editor's commands, one can create a file consisting of the representations of many pictures. Typically, we create files of pictures in

one long photographic session and then use them in subsequent experiments.

The editor also has a set of commands for digitizing video input that involves the parameters to be used in digitizing each picture. The digitization can be either single bit (indicating black or white) or the 16-level gray scale. The resolution can be either full resolution or half resolution; half resolution is useful for conserving storage space. The position of the picture can be determined by specifying the size of the margins at the top, bottom, left, and right of the picture. Finally, there is the digitize command itself, which allows the user to digitize what is coming in from the camera. We can repeatedly digitize the input until the lighting or the camera's focus is right. When satisfied with the digital representation that is displayed on a monitor, we can write that digital representation onto the file.

The file created by the editor is used to present the pictures in the course of an experiment. The file has a directory at the beginning that contains the address of each picture. This allows the pictures to be accessed in any order. During the experiment, the experimental program can request an arbitrary picture. By looking in the directory of the picture file, the program can locate the address of that picture and fetch the representation of that picture.

The picture representations, the data, are transferred using DMA from the disk to the digitizer's memory and then displayed. The approximate time to read one picture chosen at random and to display it is about .5 sec. The delay is determined by the DMA and the storage device, in our case, hard cartridge disks. The DMA routines provided by the operating system are used to transfer the data between the disk and the controller's memory. The DMA facility is available because digital controller's 16K of memory is configured as part of the processor's memory, even though the processor itself cannot generally address those locations. The controller sits on the PDP-11/04's Unibus just like the regular memory.

A picture editor allows the user to modify or create the contents of a particular picture at the keyboard. This editor has a cursor that can be moved around the video image and allows pixels to be changed from white to black or vice versa, which allows the user to clean up pictures where there is noise caused by inaccuracies in the A/D conversion. Such noise sometimes occurs because of lighting deficiencies in creating the original picture. The pixel editor can be used to draw simple line drawings consisting of squares, lines, rectangles, and some curves. Also, pictures can be transformed by rotation, size scaling, translation, and so on.

The controller also has alphanumeric capability using a standard video terminal ROM. We have the Beehive Superbee ROM, consisting of 80 columns and 22 rows. The ROM uses a 5 by 7 dot matrix to represent each character, with two scan lines below to allow for descending characters and another three scan lines for spacing between the character lines. Alphanumerics are used primarily in reading experiments. Paragraphs are stored in a regular ASCII file and then displayed one at a time using the controller's ROM. We have macros, subroutines, and system macros that resemble ordinary video terminal controlling macros. We can send a group of characters to the device and treat it as though it were a video terminal.

One somewhat novel application is a combination of alphanumeric displays with picture displays. These can be presented on the same screen as long as the two display types do not occur in the same horizontal band of video. In other words, picture and text can occur with one above the other but not next to each other. This facility is used when we want the subject's fixation coordinates displayed while he is scanning a picture. We can then test the correspondence between the coordinates the computer receives and what the subject is looking at. We also have an alphanumeric grid to help in positioning pictures and a grid for calibrating the eye fixations.

While this controller is useful and flexible, there are certain ways in which it could be improved. First of all, the A/D converter is of fairly low quality, causing errors in A/D conversion. Second, a home-built device has all the inherent maintenance, servicing, and replacement part difficulties. Third, parts cost $2,000. Fourth, there are timing limitations: One cannot present pictures with a 0-msec interstimulus interval (ISI). The minimum ISI is approximately 500 msec if a new picture representation has to be read in from the disk. The ISI can be negligible if the two representations can be stored in the controller's memory at the same time. This occurs for small pictures or paragraphs. In these cases the screen image can be changed between video fields. A final limitation is that moderately fast, large disks are necessary to get high performance from this controller. We have cartridge disks that each hold 5 megabytes, and that is enough to store approximately 160 pictures.

In summary, the digital video controller is one answer to the problem of pictorial presentation. Since the presentation is on-line, it allows for the possibility of interacting with the stimulus, for example, by altering the stimulus during the actual task. This device is fairly flexible and allows for inputting a variety of materials with relative ease. This technical project is the result of a merger of our scientific interests in image processing and the currently available technology.

REFERENCE NOTE

1. Pellegrino, J. W., & Glaser, R. *Components of inductive reasoning.* Paper presented at NOR/NPRDC Conference, San Diego, California, March 6-9, 1978.

REFERENCES

CARPENTER, P. A., & JUST, M. A. Reading comprehension as eyes see it. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension.* Hillsdale, N.J: Erlbaum, 1977.

FORBES, A. R. An item analysis of the advanced matrices. *British Journal of Psychology*, 1964, **34**, 223-236.

HUNT, E. Quote the Raven? Nevermore! In L. W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, N.J: Erlbaum, 1974.

JUST, M. A., & CARPENTER, P. A. Eye fixations and cognitive processes. *Cognitive Psychology*, 1976, **8**, 441-480. (a)

JUST, M. A., & CARPENTER, P. A. The role of eye fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation*, 1976, **8**, 139-143. (b)

POWELL, J. M. *The video I/O channel: A versatile design for real-time digitization and display of images.* Unpublished master's thesis, Carnegie-Mellon University, Department of Electrical Engineering, 1978.

RAVEN, J. C. *Advanced progressive matrices Sets I and II.* New York, N.J: The Psychological Corporation, 1974. (Originally published, 1962.)

STERNBERG, R. J. *Intelligence, information processing and analogical reasoning: The componential analysis of human abilities.* Hillsdale, N.J: Erlbaum, 1977.

NOTE

1. The programs were written by Craig Bearer, Michael Cronin, and Charles Kollar.